Capturing Transitions Between Users' Semantically Meaningful Places Using Mobile Devices

Tom Lovett
Department of Computer Science
University of Bath
Bath, UK
trl20@bath.ac.uk

Eamonn O'Neill
Department of Computer Science
University of Bath
Bath, UK
eamonn@cs.bath.ac.uk

ABSTRACT

Due to their ubiquity and ever-increasing technical capabilities, mobile devices are often used as data collection tools by researchers in multiple fields, notably HCI and Ubicomp. Although the data gathered by mobile devices can be generated from sources such as the device users, it is difficult for researchers to capture ground truth and verify data integrity beyond controlled laboratory studies. This lack of knowledge about data integrity may, in turn, affect the quality of higher-level inferences made using the data. In this paper, we report on the experience and results of a hybrid laboratory/field study in which we use mobile devices to infer the moment at which users transition between self-defined semantically meaningful personal places. The results show that filtered device motion does appear to reflect these moments of transition well, but the nature of the research question makes verification difficult in a field study.

Categories and Subject Descriptors

C.3 [**Special-purpose and Application-based Systems**]: Real-time and embedded systems

Keywords

mobile data collection, context inference, context transition, ground truth, field study

1. INTRODUCTION

The mobile device is evolving into an important tool for context aware systems. As devices such as smartphones become ever more sophisticated, with increased capabilities for sensing and computing, their utility for context awareness increases. Indeed, the mobile device offers unique data to context aware systems due to its ubiquity across diverse environments and its proximity to its users. It is a powerful enabler for sensing and computation in the real world.

As such, mobile devices are extremely useful tools for data collection in user studies. They allow researchers to gather data from users in an unobtrusive manner, which enables the capture of rich data that otherwise cannot be obtained without great effort. This is particularly apparent at the HCI level in mobile context aware computing. One notable example is semantic place recognition, where the goal of the system is to learn and recognise users' *subjective* and semantically meaningful personal places [5, 10] rather than the *objective* and absolute locations associated with location-based services (cf. [4]). The subjective element of place recognition adds complexity to sensing and inference and, as such, verification of data integrity is non-trivial.

In this paper, we focus on capturing the moments of transition between users' semantically meaningful places using mobile device motion data. Transition events are useful as they can – if captured – act as triggers for resource-intensive sensors or user notifications [6, 15]. Existing mobile place recognition systems use motion as simple triggers for higher-level sensing, but none have systematically analysed how mobile device motion relates to transitions between places.

We contribute a systematic analysis of a semantic place transition detection system using mobile device motion. More specifically, we analyse two factors – moving average time windows and weighting methods – and show that they have significant effects on place transition detection performance. We first review work related to semantic place recognition with mobile devices, before describing our approach to place transition detection with mobile device motion. We then outline the design of a hybrid laboratory/field study that captures users' natural transitions between places in addition to high-precision ground truth, before reporting the results of our analysis.

2. BACKGROUND AND RELATED WORK

In this section we provide some background to mobile semantic place recognition and contrast our contribution against relevant work in the field. The problem of semantic place recognition has received attention from researchers in recent years due to the ubiquity of mobile devices that can enable recognition in the field. There are generally two approaches to recognition: geometric-based, where spatial coordinates are used for clustering into places; and fingerprint-based, where the signatures of signals in the environment are used to identify place 'zones'.

Notable geometric-based systems include: Askbrook and Starner's GPS-based work [1], which clusters GPS coordinates *post hoc* to learn users' significant locations; Kang *et al.* [7], who use a time-based approach to cluster GPS coordinates and extract places in an *ad hoc* manner; and Liao *et*

MCSS'12, June 25, 2012, Low Wood Bay, Lake District, UK.

al. [12], who use supervised learning to identify place and activity transition sequences *post hoc*. The main disadvantage of these systems, however, is their dependency on GPS, which means they do not work well for finer-grained indoor places.

Notable work using fingerprint-based systems includes: Hightower *et al.* [5] and Kim *et al.*'s semantic place recognition using wireless RF fingerprinting [9], SensLoc [10] and Loci [8]. These later systems operate well indoors, and use device motion to trigger wireless sensing, but no analyses of how well the captured motion relates to places transitions are performed. Other place recognition systems have exploited mobile devices for a multi-modal approach, e.g. [3, 13], and motion-triggering has been exploited in positioning and fingerprinting systems, e.g. [2, 3, 11] for sensor activation and switching.

Although these systems use mobile device motion to trigger sensing and inference, none have systematically analysed how well motion relates to place transition. As such, we present such an analysis and contribute our results.

3. APPROACH

In this section we outline our approach to capturing transitions between users' semantically meaningful places using mobile device motion. First we outline the problems involved in transition detection from mobile device motion, before using them to inform our process design. We then describe our hybrid laboratory/field study in which we collect the necessary data for *post-hoc* analysis of how the factors in our design affect transition detection performance. We begin by summarising the key problems involved in capturing the moments of place transition from mobile device motion data:

1. Device motion may be a manifestation of noise or a less significant activity, e.g. the user idly playing with her device. Conversely, device motion may not always reflect user motion, e.g. the user leaving her device on a desk.

2. The intensity of motion that indicates a state of stationary or moving context may vary between contexts and between users.

3. Motion at a single point in time, or over a short period of time, may not provide enough information about whether the user is in a state of static or dynamic context. Conversely, increasing the amount of historical data to be considered could affect decision latency.

These problems provide a rationale for our process of detecting users' place transitions from mobile device motion data. The goal of the process is – in real time – to send a message when the user is transitioning into (entering) or out of (exiting) a place. Figure 1 shows the components involved in this process, and the following subsection describes its design.

3.1 System Design

The transition capture process is designed to operate on-device and uses two components commonly found in mobile motion detection systems: a binary classifier and moving average time window.

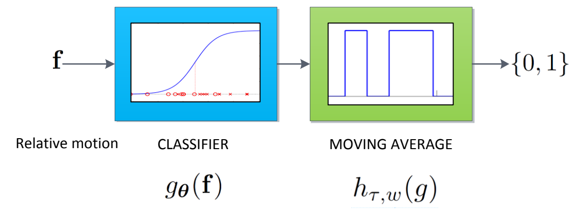

Figure 1: The process design

- **A binary classifier** addresses problem 2. It is very unlikely for motion intensity to be consistent both within and between places, and motion patterns will vary between users and the device's on-body location. The classifier outputs values in $[0, 1]$ that represent the probability of the device undergoing significant motion (or the complement probability of insignificant motion) at timestep k, given the motion intensity:

$$g_{\boldsymbol{\theta}}(\mathbf{f}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}}} \tag{1}$$

  Where $\mathbf{f}$ is the relative motion vector at k and $\mathbf{z}$ is a column vector in $\mathbb{R}^2$, defined:

$$\mathbf{z} = (1, \|\mathbf{f}\|)^{\mathrm{T}} \tag{2}$$

  Here, $\boldsymbol{\theta}$ is a parameter vector in $\mathbb{R}^2$ that controls the classifier threshold. We can learn these parameters for a specific user by training the classifier on a sample of the user's motion data in various context states.

- **A moving average filter** addresses problems 1 and 3. To minimise the effect of transient and unimportant motion, we can smooth the classifier outputs over a fixed time window, τ, so that only sustained motion can trigger a transition. This uses historical data and, as such, requires a necessary lag to operate. In addition to varying τ, we can use weighting methods, w, for the historical data, which can vary the influence from more recently acquired data. The moving average outputs high if the weighted average of the classifier outputs in τ is ≥ 0.5 and low otherwise.

3.2 Study Design

In order to evaluate system performance, we must design a study that will allow us to capture the data required for analysis in the most natural way possible. We have two requirements: naturalism, i.e. capturing data in a natural environment; and integrity, i.e. high-precision ground truth data. A field study would satisfy the requirement for naturalism, and a laboratory study would satisfy the requirement for data integrity, but neither can easily satisfy both. In the case of capturing place transitions, a hybrid study satisfies the trade-off between the two requirements.

We therefore conducted an empirical study of mobile users in a hybrid laboratory/field study. Participants were asked to visit a set of semantically meaningful personal places in a natural order, whilst being shadowed by a researcher recording ground truth. We recruited 14 participants (11 male and 3 female; aged 20-38, mean age 27) from three different daily environments: an office (7 participants), a university (6 participants), and a town centre (1 participant). In a pre-study interview, we asked them to describe their typical day's activities chronologically through transitions between semantically meaningful personal places within their environment. Immediately following the interview, we asked them to choose a sequence of these places (and activities) that that could be performed as a scripted tour. Each participant was equipped with an Android mobile device containing an accelerometer, from which the output was continually logged at $\approx$ 16Hz throughout the study. The participants underwent a training session to train their logistic regression parameters, during which they were asked to perform example within-place and between-place activities, e.g. sitting at a desk or walking, while carrying the device in a pocket or a bag. Each training session lasted 30 seconds per activity type.

The participants were then asked to undergo their previously identified place sequences and perform their previously identified example activities in each place whilst carrying the mobile device in exactly the same manner as they would at and between each place. A researcher shadowed each participant and recorded the timestamp for the transition points into and out of each place – notified orally to the researcher by the participant themselves. Due to the difficulties involved in collecting such fine-grained data over an extended period of time, the participants were asked to perform shortened versions of activities in each place, e.g. "working at desk", which would typically last for 1–2 hours, was shortened to 5–10 minutes. The transitions between places were not shortened.

3.3 Analysis

Upon completion of the hybrid study, we analysed the data in order to observe the effects of each factor on the transition detection process. The factors are w, the weighting method for the moving average filter; and τ, the time window for the moving average filter. Although the data was *logged* at 16Hz, we *sample* from it at 1Hz, so the difference between timesteps k is constant at 1 second. For the participant-specific logistic regression parameters, θ_i, we used each participant's training data to find the maximum likelihood parameters – $\hat{\boldsymbol{\theta}}_i$ – for that participant i. Once found, $\boldsymbol{\theta}_i$ was held fixed at $\hat{\boldsymbol{\theta}}_i$ for each participant i during analysis.

We chose three moving average weighting methods, w, to analyse:(i) the simple moving average (SMA), where all classifier outputs in time window τ are given equal weight; (ii) the weighted moving average (WMA), where the classifier outputs are weighted linearly over τ (with more weight given to recent motion); and (iii) the exponential moving average (EMA), where the classifier outputs are weighted exponentially over τ as follows, $\frac{2}{T_n+1}, T_n \leq \tau$, where T_n is the time-lag between the current timestep k and timestep n. As a benchmark, we also tested the process with no moving average. Finally, we evaluated 9 time windows τ at intervals of 5–10 seconds over 5–60 seconds.

3.4 Performance Measure

To measure the performance of each design, we use the precision, recall and F1 accuracy scores which account for true positive (tp), false positive (fp) and false negative (fn) classifications. Their descriptors are as follows:

- A true positive (tp) occurs when the process classifies a correct transition point according to ground truth, i.e. a place entrance or exit transition. This must be made within an acceptable time window from the ground truth transition point, accounting for moving average lag τ.

- A false positive (fp) occurs when the process classifies an incorrect transition point according to ground truth, i.e. classifying a transition outside a viable ground truth transition.

- A false negative (fn) occurs when the process fails to classify a transition point according to ground truth, i.e. not classifying a transition at the time of a viable ground truth transition.

To account for small deviations between the researcher-recorded ground truth and the exact moment of context

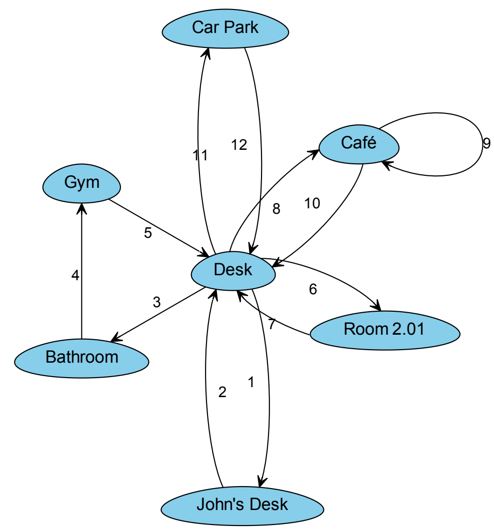

Figure 2: Graph of the place transition sequence for Participant 7. Edge labels represent the transition ordering, and each edge represents an exit and entrance transition point.

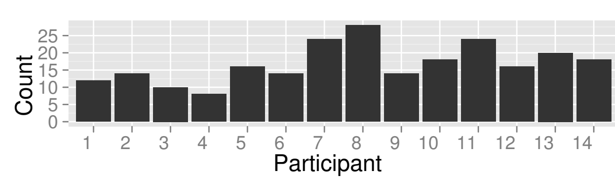

Figure 3: Transition point count distribution over the participants.

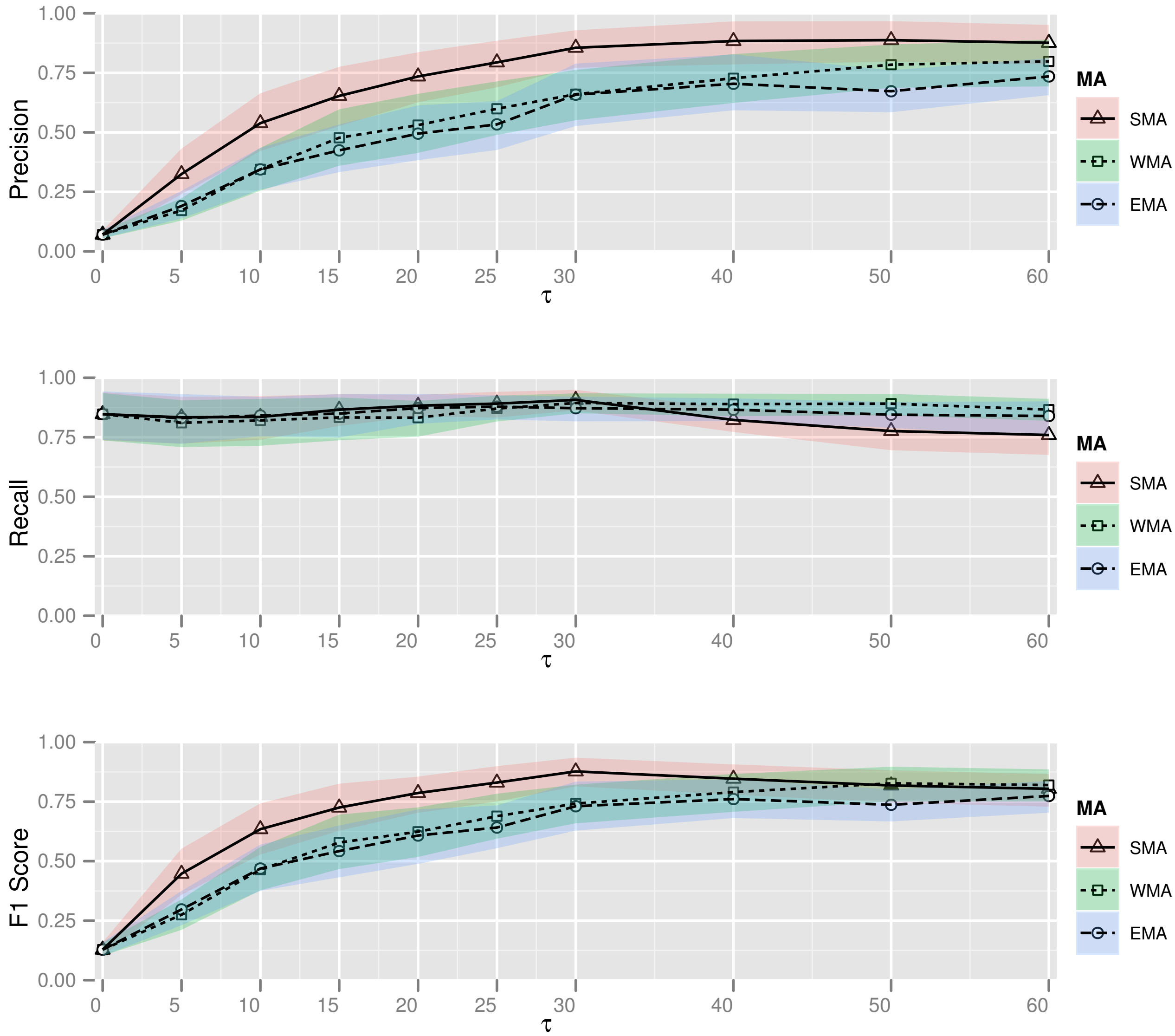

Figure 4: Plots showing the mean precision (top), recall (middle) and F1 score (bottom) for each moving average (MA) type over the time window τ. (95% confidence intervals are displayed.)

state transition, a ground truth was considered viable for 5 seconds either side of its recorded timestamp.

4. RESULTS

For the study, the environment for participants 1-6 was a university campus; 7-13 was an office; and for 14 it was a town centre. The median number of transition points for the 14 participants was 16. Common place labels included: "desk"; "café"; "canteen"; "lecture hall"; "car park"; "lab"; "gym" and "meeting". Common within-place activities included: "working", "eating", "reading" and "relaxing". Common between-place activities included: "walking" (all participants); "cycling" (participant 4); and "driving" (participant 14). Figure 2 shows a graph representation of Participant 7's place transition sequence, with each edge corresponding to two transition points (entrance and exit). Figure 3 shows the transition point distribution over each of the 14 participants.

A two-way, within subject analysis of variance (ANOVA) over the factors w and τ (and their interaction) shows that neither have a significant effect on each participant's true positive tp count. The same ANOVA shows that τ has a highly significant effect on each participant's false positive fp count ($p < 0.0001$), and w has a significant effect on the per-participant fp count ($p < 0.05$). There is also a significant interaction effect on the per-participant fp count from τ and w ($p < 0.05$). The same ANOVA shows that neither factor τ or w (nor their interaction) have a significant effect on the per-participant false negative fn count.

By encoding the tp, fp and fn counts into between-participant comparable statistics – precision, recall and F1 score – we can further analyse the effects of each factor on per-

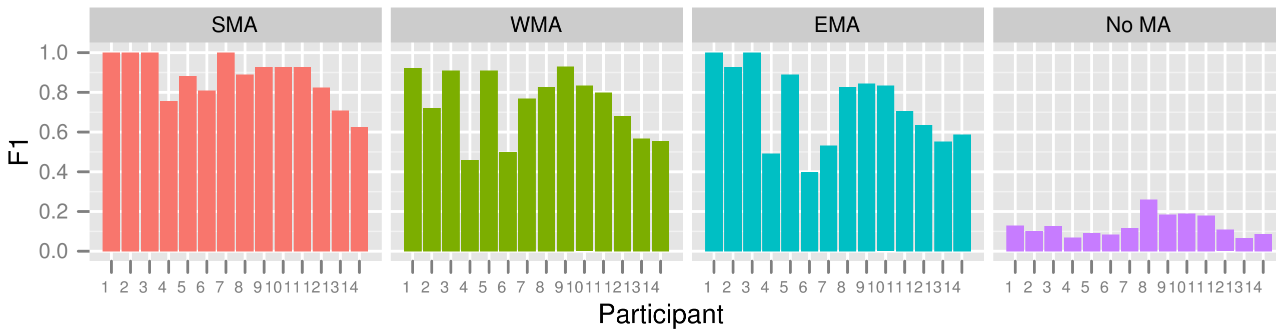

Figure 5: F1 accuracy, for each moving average factor, over the participants at $\tau = 30$s

formance. As the distributions of these statistics over the participants are unknown (and not well modelled using a normal distribution), we use a non-parametric bootstrapping method with 1000 replicates to estimate the mean and percentile confidence intervals of each statistic over the factors. Figure 4 shows the mean precision, recall and F1 scores for each w over τ. There is a significant observed improvement in precision and F1 score from no moving average by all levels of w for $\tau > 5$s ($p < 0.05$). There is a significant observed improvement in F1 score ($p < 0.05$) at $\tau = 30$s from $\tau < 15$s for the SMA; at $\tau = 50$s from $\tau < 25$s for the WMA; and at $\tau = 60$s from $\tau < 20$s for the EMA.

Figure 5 shows a more detailed overview of the F1 score distribution for the participants, partitioned by w, at $\tau = 30$s.

5. DISCUSSION

Here we discuss the observations, limitations and implications of our approach and results. The results suggest that the majority of users' significant place transitions can be captured through inference of sustained motion using a simple binary classifier and moving average window running on a mobile device. Furthermore, performance is good in a fine-grained environment, i.e. room/building level, with no specialist hardware. The F1 score of 0.87 from the best process design ($\tau = 30$s, w =SMA), and the significant improvement in observed performance from using a moving average filter, shows that mobile device motion over time is a good indicator of when a user transitions between semantically meaningful places.

5.1 Observations

The significant improvement from the absent case by all moving average types w in Figures 4 and 5 shows that smoothing transient motion and requiring *sustained motion over time* is an effective method of detecting user context state transitions. The results from the ANOVAs show that the number of per-participant true positives tp are invariant to both τ and w, showing that without smoothing, the interesting motion data, i.e. the transition points, are always present but hidden in the noisy motion generated by latent sources, e.g. users idly playing with their devices.

Interestingly, there is a peak in F1 score performance for the SMA at $\tau = 30$s (see Figure 4). This is due to an increase in false negatives fn and consequent decrease in recall as the time window τ increases beyond the shorter transitions for many participants, e.g. walking from a desk to a meeting which may take less than 30s. The peak is more apparent earlier for the SMA due to its unbiased weighting over the entire time window. The slight improvement using the SMA rather than the WMA or EMA shows that equal weighting of data in τ – rather than biasing toward recency – is likely to be the superior choice, not least because of the improvement in classification latency (compare the approximately equal performance in Figure 4 of the SMA at $\tau = 20$s to the WMA and EMA at $\tau = 40$s).

Clearly the greatest improvement comes from reducing the per-participant false positive fp count. Aside from in-place idle motion (e.g. from the device in a pocket, or the user idly playing with it), these were generally caused by participants undergoing periods of "start-stop" motion both within and between places, e.g. participant 12 using their device for a phone call; participant 6 moving within a large shop; participant 4 cycling; and participant 14 driving. A few false negatives fn were caused by participants leaving their device at their desk to travel to a nearby location, e.g. a printer (participant 7), or undergoing short transitions (with duration $< \tau$), e.g. stopping to talk to a colleague *en route* to another location (participant 13).

5.1.1 Hybrid Study

Notable observations from our hybrid study approach included the lack of cognitive overload for the participants. With the shadow monitoring them, some participants noted that they didn't have to "stop and think" (participant 13) about writing something down, or what the should be doing next, although some participants noted that – as well as the presence of the shadow – the time shortening during places felt a little artificial, even with them performing their natural activities in each. Another notable observation was the participants' willingness to undertake the study under the hybrid conditions: the majority said that – for privacy reasons – they would not undertake the study if the shadow (or other observation device, e.g. a camera) were to be present throughout their entire day, i.e. a full observational field study.

5.2 Limitations

One of the key limitations of the study were the environments. We were focused on 'local' environments – offices, campuses and a town centre – so we cannot easily generalise the performance from these results to multiple, more global, environments. Early indications of how the detection process deals with vehicular motion, i.e. participant 14, suggest that the stop-start nature of driving will impact on performance due to the fixed threshold of the trained classifier. However, using multiple fixed or adaptable thresholds may alleviate these problems. Furthermore, data fusion with other sources of context data that suggest, for example, that the user is in a vehicle, e.g. in-car Bluetooth, could improve performance in these situations, as could incorporating others' work into detecting transport types from mobile device motion, e.g. [14].

There are also limitations with the ecological validity of the hybrid field study. First, although the participants were asked to carry their mobile device in a naturalistic manner, e.g. in a pocket or bag, we could not capture entirely realistic idle motion profiles due to the shortening of the context periods. Furthermore, the performance of the participants' activities was necessarily artificial, i.e. they were enacted for the purpose of the study rather than to achieve a specific goal which could, in turn, affect the ecological validity of the captured motion data.

5.3 Implications

The output of this work results in a lightweight mobile service that can report genuine place transitions to any application that requires them. This has implications for various applications that focus on semantic place recognition, e.g. [10], notification delivery systems, e.g. email or SMS, or in-situ user prompting. The results and feedback from using the hybrid study approach show that it (the approach) can be used to acquire useful results that could not otherwise be obtained reliably through laboratory or field studies.

6. CONCLUSION

In this paper, we have shown how transitions between users' semantically meaningful places can be captured using mobile devices – specifically through motion. Although thresholds and moving averages have previously been used as indicators of context change, no one has yet systematically analysed the effects on performance when their parameters are varied; particularly in the case of capturing transitions between semantically meaningful personal places. We have shown – through a hybrid laboratory/field study in which we capture high-precision ground truth – that both the moving average weighting method and time window have significant effects on the number of per-participant false positive transitions, and results suggest that a simple moving average (SMA) with a ≈ 30s time window is an appropriate choice for good transition capture performance.

For future work, we plan to use the transition detection method as an enabler to trigger alerts and notifications in further field studies involving mobile devices.

7. ACKNOWLEDGEMENTS

Eamonn O'Neill's research is supported by a Royal Society Industry Fellowship at Vodafone Group R&D. We also wish to thank our study participants for their time and efforts.

8. REFERENCES

[1] D. Ashbrook and T. Starner. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5):275–286, Oct. 2003.

[2] P. Bolliger, K. Partridge, M. Chu, and M. Langheinrich. Improving Location Fingerprinting through Motion Detection and Asynchronous Interval Labeling. In T. Choudhury, A. J. Quigley, T. Strang, and K. Suginuma, editors, *LoCA*, pages 37–51. Springer, 2009.

[3] Y. Chon, E. Talipov, and H. Cha. Autonomous Management of Everyday Places for a Personalized Location Provider. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews,.*, (99):1–14, 2011.

[4] S. Harrison and P. Dourish. Re-Place-ing Space : The Roles of Place and Space in Collaborative Systems. In *Proc. CSCW '96*, pages 64–76, 1996.

[5] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes. Learning and recognizing the places we go. In *Proc. UbiComp '05*, pages 903–903, 2005.

[6] J. Ho and S. S. Intille. Using Context-aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In *Proc. CHI '05*, pages 909–918, 2005.

[7] J. H. Kang, W. Welbourne, B. Stewart, and G. Borriello. Extracting places from traces of locations. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(3):58, July 2005.

[8] D. H. Kim, K. Han, and D. Estrin. Employing User Feedback for Semantic Location Services. In *Proc. Ubicomp '11*, pages 217–226, 2011.

[9] D. H. Kim, J. Hightower, R. Govindan, and D. Estrin. Discovering Semantically Meaningful Places from Pervasive RF-Beacons. In *Proc. UbiComp '09*, pages 21–30, 2009.

[10] D. H. Kim, Y. Kim, D. Estrin, and M. B. Srivastava. SensLoc : Sensing Everyday Places and Paths using Less Energy. In *Proc. SenSys '10*, pages 43–56, 2010.

[11] M. B. Kjærgaard, J. Langdal, T. Godsk, and T. Toftkjær. EnTracked : Energy-Efficient Robust Position Tracking for Mobile Devices. In *Proc. MobiSys '09*, pages 221–234, 2009.

[12] L. Liao, D. Fox, and H. Kautz. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields. *The International Journal of Robotics Research*, 26(1):119–134, Jan. 2007.

[13] R. Montoliu and D. Gatica-Perez. Discovering Human Places of Interest from Multimodal Mobile Phone Data. In *Proc. MUM '10*, pages 1–10, 2010.

[14] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava. Using Mobile Phones to Determine Transportation Modes. *ACM Transactions on Sensor Networks*, 6(2):1–27, Feb. 2010.

[15] T. Sohn, K. A. Li, G. Lee, I. Smith, J. Scott, and W. G. Griswold. Place-Its : A Study of Location-Based Reminders on Mobile Phones. In *Proc. Ubicomp '05*, pages 232–250, 2005.